

# Prediction and clinical utility of a liver cancer risk model in Chinese adults: a prospective cohort study of 0.5 million people

Chengxiao Yu<sup>1,2†</sup>, Ci Song<sup>1,2†</sup>, Jun Lv<sup>3,4†</sup>, Meng Zhu<sup>1,2</sup>, Canqing Yu<sup>3</sup>, Yu Guo<sup>5</sup>, Ling Yang<sup>6</sup>, Yiping Chen<sup>6</sup>, Zhengming Chen<sup>6</sup>, Tao Jiang<sup>1,2</sup>, Hongxia Ma<sup>1,2</sup>, Guangfu Jin<sup>1,2</sup>, Hongbing Shen<sup>1,2</sup>, Zhibin Hu<sup>1,2\*</sup>, Liming Li<sup>3\*</sup>, on behalf of the China Kadoorie Biobank Collaborative Group<sup>††</sup>

<sup>1</sup> Department of Epidemiology, China International Cooperation Center on Environment and Human Health, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing 211166, China

<sup>2</sup> Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Personalized Medicine, Nanjing Medical University, Nanjing 211166, China

<sup>3</sup> Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing 100191, China

<sup>4</sup> Key Laboratory of Molecular Cardiovascular Sciences (Peking University), Ministry of Education, Beijing 100191, China

<sup>5</sup> Chinese Academy of Medical Sciences, Beijing 100730, China

<sup>6</sup> Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: [10.1002/ijc.33487](https://doi.org/10.1002/ijc.33487)

<sup>†</sup> These authors contributed equally to this work.

\* **Correspondence to:** Zhibin Hu, Department of Epidemiology, Center for Global Health, School of Public Health, Nanjing Medical University, 101 Longmian Avenue, Nanjing 21116, China, Tel: +86-25-86868437, E-mail: zhibin\_hu@njmu.edu.cn; Liming Li, Department of Epidemiology and Biostatistics, Peking University Health Science Center, 38 Xueyuan Road, Beijing 100191, China, Tel: +86-10-82801528, Email: lmlee@vip.163.com.

<sup>††</sup> The members of steering committee and collaborative group are listed in the online-only supplemental material.

**Abbreviation:** AFP, alpha fetoprotein; ALT, alanine aminotransferase; BMI, body mass index; CI, confidence interval; DCA, decision curve analysis; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B virus surface antigen; HBV, hepatitis B virus; ICD-10, the 10th revision of the International Classification of Diseases; MET, metabolic equivalent of task; ROC, receiver operating characteristic curve;

### **Novelty and Impact**

The personalized risk prediction model for liver cancer based on HBV infection status, medical history, metabolic risk factors and lifestyle-related factors from nationwide prospective cohort study is lacking. Based on these routinely available predictor factors, we established a personalized risk prediction model. Decision curve analysis revealed that use of the model in selecting participants for screening improved benefit at threshold probabilities of 2% 10-year risk, compared with current guideline of screening all HBsAg carriers.

**Abstract**

China has made rapid progress in reducing the incidence of HBV infection in the past three decades, along with a rapidly changing lifestyle and aging population. We aimed to develop and validate an up-to-date liver cancer risk prediction model with routinely available predictors and evaluate its applicability for screening guidance. Using data from the China Kadoorie Biobank, we included 486,285 participants in this analysis. Fifteen risk factors were included in the model. Flexible parametric survival models were used to estimate the 10-year absolute risk of liver cancer. Decision curve analysis was performed to evaluate the net benefit of the model to quantify clinical utility. A total of 2,706 participants occurred liver cancer over the 4,814,320 person-years of follow-up. Excellent discrimination of the model was observed in both development and validation datasets, with c-statistics (95% CI) of 0.80 (0.79 to 0.81) and 0.80 (0.78-0.82) respectively, as well as excellent calibration of observed and predicted risks. Decision curve analysis revealed that use of the model in selecting participants for screening improved benefit at a threshold of 2% 10-year risk, compared with current guideline of screening all HBsAg carriers. Our model was more sensitive than current guideline for cancer screening (28.17% vs. 25.96%). We developed and validated a CKB-PLR (Prediction for Liver cancer Risk based on the China Kadoorie Biobank study) model to predict the absolute risk of liver cancer for both HBsAg seropositive and seronegative populations. Application of the model is beneficial for precisely identifying the high-risk groups among the general population.

**Key words:** Liver cancer; Risk prediction model; Cohort study.

## Introduction

China possesses the highest incidence of liver cancer globally and accounts for almost half of the liver cancer burden in the world.<sup>1, 2</sup> The prognosis of liver cancer patients is generally poor, with a 5-year survival rate of less than 5%.<sup>3</sup> However, prognosis varies strongly by disease stage at diagnosis. Early-stage liver cancer is amenable to curative therapy, enabling a 5-year survival rate of 40%-70%.<sup>4, 5</sup> Taken together, these facts highlight the importance of screening and early diagnosis of liver cancer.

In China, hepatitis B virus (HBV) infection has been implicated in the etiology of 60-80% of liver cancer.<sup>6, 7</sup> Hence the current recommended screening approach is limited in HBV carriers.<sup>5</sup> Due to the implementation of maternal-to-infant blocks and high vaccination coverage among children,<sup>8, 9</sup> China has made rapid progress in reducing the incidence of HBV infection in the past three decades.<sup>10, 11</sup> The prevalence rate of hepatitis B surface antigen (HBsAg) declined from 9.8% in 1992 to 6.1% as an estimated prevalence in 2016.<sup>10, 12-14</sup>

Meanwhile, a rapid shift in lifestyle and aging population have led to dramatic changes in the risk factor pattern for liver cancer in China during the past decades.<sup>6, 13</sup> In addition, the prevalence of metabolic diseases centered on insulin resistance is increasing, and a growing number of cases of metabolic diseases have been observed transformed to liver cancer.<sup>15-17</sup> Accordingly, maintaining a healthy lifestyle (i.e., smoking cessation, drinking cessation, moderate blood glucose levels, exercise and moderate weight) can dramatically reduce the risk of developing liver cancer and dying from liver-related causes.<sup>6, 13</sup> However, previous liver cancer prediction models were limited by focusing on specific disease populations (e.g., HBV infection, diabetes)<sup>18, 19</sup> or putting excessive weight on clinical indicators (e.g., alanine aminotransferase (ALT), AFP).<sup>20</sup> Thus, a new personalized risk prediction model for liver cancer based on routinely available predictors such as HBV infection status, medical history, metabolic risk factors and lifestyle factors is urgently needed to encourage high-risk people to be screened early.

Taking advantage of the largest nationwide prospective cohort of 0.5 million in China, we conducted the CKB-PLR (Prediction for Liver cancer Risk based on the China Kadoorie Biobank study) project to establish and validate the personalized Chinese 10-year liver cancer risk prediction model. We further evaluated the potential clinical benefits of the liver cancer model.

## Methods

### *Study population*

The CKB is a prospective cohort study of 512,714 adults aged between 30 and 79 years at the time of recruitment. The cohort was conducted between 2004 and 2008 in ten study areas geographically (5 urban and 5 rural) spread across China. Details of the cohort design, survey methods, and population characteristics have been described previously.<sup>21</sup> The study was approved by the Ethical Review Committee of the Chinese Center for Disease Control and Prevention (Beijing, China) and the Oxford Tropical Research Ethics Committee, University of Oxford (UK).

In the present analysis, we excluded participants who reported a medical history of liver cancer at baseline (n=37). Participants with missing covariate data were also excluded, e.g., parental or sibling history of cancer (n=14,831), body mass index (n=2), HBV surface antigen (n=11,731) and random plasma glucose level (n=8,340), leaving 486,285 participants for the CKB-PLR model analysis. Using the randomized grouping method, a development/validation split where the model was fitted to 80% (n=389,028) of the data and evaluated on the remaining 20% (n=97,257) (**Supplementary Figure 1**).

### *Variable selection*

In the baseline questionnaire, participants completed an interviewer-administered, laptop-based questionnaire. We focused on established predictor risk factors for liver cancer that are routinely available or easily ascertained by general practitioners during a standard consultation. Ultimately, baseline information in the CKB-PLR project included the following: (i) sociodemographic characteristics including age (years), gender (male or female), residential area (urban or rural), and educational level (primary school and below, or middle school and higher); (ii) lifestyle behaviors including tobacco consumption (never and occasional smoker, or ever smoker including ex- and current regular smoker), alcohol consumption<sup>22</sup> (never weekly drinker, or ever weekly drinker including ex-regular, reduced-intake, or current regular drinker), and physical activity (metabolic equivalent of task [MET]-h/day); (iii) personal medical history (yes or no), including physician-diagnosed cancer, cirrhosis or chronic hepatitis, gallstone or gallbladder disease, and diabetes; (iv) family medical history (parental or siblings history of cancer; yes or no); (v) physical measurement including body mass index (BMI; kg/m<sup>2</sup>); (vi) blood test results including random

plasma glucose (mmol/L) and hepatitis B surface antigen (positive or negative). Further details of the physical measurement and blood test results are described in **Supplementary Text 1**.

### ***Follow-up of the incidence of death and liver cancer***

Study participants were followed up for liver cancer morbidity and mortality information, mainly through existing disease monitoring systems.<sup>21</sup> The vital status of each participant was determined periodically through the local Disease Surveillance Points system,<sup>23</sup> supplemented by regular checks against local residential records and by annual active confirmation through street committees or village administrators. Additional information about major diseases and any episodes of hospitalization was collected through linkages, via each participant's unique national identification number, with disease registries and national health insurance claims databases, which have almost universal coverage in the study areas. Trained staff, blinded to the baseline information,<sup>21</sup> coded all cases with the 10th revision of the International Classification of Diseases (ICD-10). In the present study, participants were followed up from the date of baseline enrollment until the date of diagnosis of liver cancer (ICD-10 code: C22), date of death, or January 1, 2017, whichever occurred first.

### ***Statistical analysis***

All statistical analyses were conducted using R (version 3.5.0). Poisson regression is often used in the analysis of the incidence of rare outcomes in the cohort study, and can be used to estimate adjusted incidence rates.<sup>24</sup> We used Poisson regression models to estimate age-, sex-, and residential area-adjusted incidence rates of liver cancer events per 100,000 person-years (**Table 1; Supplementary Table 1**).

### ***Model development and validation***

For model development, we separately modeled the hazard of liver cancer and the hazard of death from all causes to account for death as competing events. Flexible parametric survival models were used on the cumulative hazard scale to estimate baseline hazards and hazard ratios.<sup>25</sup> We used restricted cubic splines to model potential nonlinear relationships with continuous variables.<sup>26</sup> Age was modeled by a restricted cubic spline with internal knots at 50, 60, and 70 years, and the random plasma glucose was fitted with boundary knots at 2.8 and 11.1 with internal knots at 5.6 and 6.8. Nonproportional hazards were investigated by fitting interactions between covariates and the timescale. Separate models for liver cancer risk were fitted for HBsAg seronegative and HBsAg seropositive participants due to the significant

variance in liver cancer incidences. The model for risk of death included age at baseline, sex, residential area, education previous cancer diagnosis, smoking status, and alcohol consumption as covariates. We estimated the 10-year absolute risk of liver cancer for each participant using the cumulative incidence function.<sup>27</sup> An easy-to-use R code for calculating the model-predicted 10-year absolute probability of liver cancer is available from GitHub (<https://github.com/meditation-NMU/CKB-PLR>).

We performed an internal validation of the development model. The performance of the model was assessed by calibration, which measures the agreement between observed and predicted liver cancer risk, and by discrimination ability to differentiate between participants who experienced liver cancer and those who did not. Plotting the mean predicted probability against the mean observed probability of liver cancer by a tenth of the predicted absolute risk was used to assess the calibration of the model. Model discrimination was assessed with the concordance statistic (c-statistic). To evaluate the prediction efficiency of the model in the general population, we compared the performance of the model with that of the current HBsAg seropositive strategy (recommending liver cancer screening for all HBsAg seropositive participants).<sup>5</sup>

### *Clinical utility*

Decision curve analysis (DCA) was used to evaluate the clinical utility of the prediction model.<sup>28,29</sup> Such a decision may apply to more or less intensive screening. The key part of the DCA is the net benefit, which is the number of true-positive classifications (in this example, the benefit of liver cancer screening to a participant who would have developed liver cancer) minus the number of false-positive classifications (in this example, the harm of unnecessary liver cancer screening in a participant who would not have developed liver cancer). The false positives are weighted by a factor related to the relative harm of a missed liver cancer versus an unnecessary liver cancer screening. The weighting is derived from the threshold probability to develop liver cancer using a defined landmark time point (e.g., liver cancer risk at 10 years).<sup>30</sup> Decision curves plot the predicted net benefit of the risk prediction model versus risk thresholds (**Supplementary Text 2**). A 10-year risk of ~2% is considered to stratify participants as high-risk and available in cancer screening,<sup>31,32</sup> and this is also the threshold risk for liver cancer.<sup>33</sup> Furthermore, in order to assess the clinical usefulness of our prediction model around a threshold of 2% absolute risk, the decision curve analysis also showed the net benefit of the

CKB-PLR model and all other alternatives across all possible risk thresholds

## Results

### *Distribution of participant characteristics*

Among the 486,285 study participants, there were 2,706 incident liver cancer diagnoses over 4,814,320 person-years of follow-up (median [interquartile range] length of follow-up, 10.12 [9.19-11.09] years). Demographic characteristics and liver cancer risk factors are provided in **Table 1**. Patients diagnosed with liver cancer were older, more likely to be male, had a lower educational level, had a family history of cancer, were more likely to be regular smokers or alcohol drinkers, were more likely to be underweight, and had less physical activity than participants without liver cancer ( $P < 0.001$ ). As expected, HBV infection, history of cirrhosis or chronic hepatitis, history of gallstone or gallbladder disease, and diabetes (or with a higher random plasma glucose) also differed significantly among participants with and without liver cancer ( $P < 0.001$ ).

### *Development of the CKB-PLR model*

In the development and validation datasets, with a consistent proportion of HBsAg seropositive participants, similar incidence rates of liver cancer were observed among the two datasets in different HBsAg statuses (**Supplementary Table 1**).

The hazard ratios and 95% confidence intervals (CIs) for the CKB-PLR model based on the development dataset are listed in **Table 2**. As expected, differences in liver cancer risk were observed according to the categories of liver cancer predictive factors. In the multivariable model, age and random plasma glucose levels had a significant nonlinear association with liver cancer risk ( $P < 0.05$  for nonlinearity) (**Supplementary Figure 2**). The prediction performance of the CKB-PLR model based on the development dataset is shown in **Supplementary Figure 3**. The observed 10-year risk of liver cancer by a tenth of the CKB-PLR model-predicted risk, and the calibration plots suggest excellent calibration agreement of observed probabilities with predicted probabilities of developing liver cancer in 10 years (**Supplementary Figure 3A**). The receiver operating characteristic curve (ROC) for the development model is shown in **Supplementary Figure 3B**. When applied to all participants regardless of HBsAg status, the CKB-PLR model was able to discriminate well between participants who were diagnosed with liver cancer and those who did not within 10 years (c-statistic [95% CI] = 0.80 [0.79 to 0.81]). The



discriminative ability performed equally well when applied only to HBsAg seronegative (c-statistic [95% CI] = 0.76 [0.75 to 0.77]) or seropositive (c-statistic [95% CI] = 0.76 [0.74 to 0.77]) participants.

### ***Performance of the CKB-PLR model***

When the model was applied to the validation dataset, a higher distribution of the 10-year predicted probability was observed in participants who were diagnosed with liver cancer compared to those who were not within the first 10-years (**Figure 1A**). The distributions of 10-year predicted risk by age stratification are presented in **Figure 1B**. As expected, a gradient distribution for 10-year predicted risk of liver cancer was observed across age categories, which can be interpreted as advancing age associated with increased risk for liver cancer and a greater burden of risk factors with aging. Subsequently, the calibration and discriminative ability of the model were further validated in the validation dataset. Similar to the development model, the observed probabilities coincided with the predicted probabilities of developing liver cancer in 10 years in the validation dataset (**Figure 2A**). **Figure 2B** shows the ROC curve for the fitted model, and the sensitivity and specificity of the model are plotted by the predicted probability cutoff in **Figure 2C**. The c-statistic of the model applied to the validation dataset was 0.76 (95% CI = 0.73-0.78) for HBsAg seronegative participants, 0.76 (95% CI = 0.73-0.80) for HBsAg seropositive participants, and 0.80 (95% CI = 0.78-0.82) for all participants. Overall, internal independent validation suggested good performance of the CKB-PLR model when applied to new observations or data sets, due to the high capability of calibration and discrimination. As the most dominant subtype of liver cancer, hepatocellular carcinoma (HCC) was used as the outcome in the subgroup analysis. Our prediction model analysis also yielded similar good calibration and discriminative ability (**Supplementary Table 2; Supplementary Figure 4**).

To assess the possible implications of adopting a model-based inclusion criterion for liver cancer screening, we compared the sensitivity and specificity of the CKB-PLR model-based strategy with that of the HBsAg seropositive criteria in the validation dataset. All 3,069 HBsAg seropositive participants were eligible for screening in the validation dataset when adopting the HBsAg seropositive criteria. For an equal number of participants determined by the CKB-PLR model, participants with a 10-year predicted risk higher than 1.926% were eligible. Applying the HBsAg seropositive criteria to the validation dataset would yield a specificity of 96.96% with

a sensitivity of 25.96%. It yields a higher validity when setting the probability cut point to 1.926% using the CKB-PLR model-based strategy, with a specificity of 96.97% and sensitivity of 28.57% (**Supplementary Table 3**). **Supplementary Figure 5** shows the distribution of predicted probabilities by the HBsAg seropositive criteria eligibility and liver cancer event status at the 10-year follow-up. A vertical line is drawn at 1.926%, 68% of HBsAg seropositive participants have predicted risks exceeding this threshold, and the CKB-PLR model criteria reduced the likelihood of false positive and false negative.

#### *Assessment of potential clinical benefits for the CKB-PLR model*

A comprehensive overview of the net benefit for a range of harm-benefit thresholds at 10-year liver cancer risk is provided by the decision curves (the general CKB participants, **Figure 3**). Based on the probability threshold, the decision analysis curve was leveraged to evaluate the clinical application of a prediction model. **Table 3** shows the CKB-PLR model's sensitivity, specificity, and net benefit at threshold of 2% predicted absolute risk. In the comparison of using the CKB-PLR model (a threshold of 2% 10-year absolute risk) with HBsAg seropositive strategy for selection of participants for screening, the sensitivities were 28.17% versus 25.96%, the specificities were 97.12% versus 96.96% for liver cancer detection within 10 years. And the CKB-PLR model had greater net benefit than HBsAg seropositive strategy at a risk threshold of 2% (0.000912 (0.000805-0.001022) versus 0.000734 (0.000623-0.000826)) (**Table 3; Figure 3**). **Figure 3** compares the net benefits of using the CKB-PLR model for stratifying participants for screening with those of other alternatives, including the HBsAg seropositive strategy (recommending liver cancer screening for all HBsAg seropositive participants) and the extreme strategies of screening everyone or no one. This analysis showed that the ability of CKB-PLR model to predict liver cancer provided a better benefit than all other alternatives across 10-year absolute risk thresholds from 0.5% to 5%. In this range, the CKB-PLR model offers the highest net benefit. Furthermore, the decision curve analysis for preventive liver cancer screening showed the potential clinical utility of the CKB-PLR model between thresholds of 0.5-5% 10-year liver cancer risk for HBsAg seronegative ('opt-in' policy) and seropositive ('opt-out' policy) participants (**Supplementary Figure 6; Supplementary Table 4**).

## **Discussion**

Utilizing this large prospective cohort of 0.5 million adults in China, we developed a model, the CKB-PLR, based on the most known and routinely available risk factors for liver cancer. By assessing participants at substantially different levels of 10-year absolute liver cancer risk, the model had high calibration and discriminative ability. In addition, the clinical utility assessment of the CKB-PLR model showed an improvement in risk counseling.

Because of the high prevalence and critical role of HBV in the development of liver cancer,<sup>34</sup> there have been multiple HBV-associated liver cancer risk prediction score systems, including REACH-B,<sup>18</sup> CU-HCC,<sup>35</sup> and GAG-HCC.<sup>36</sup> These scoring systems incorporated demographic characteristics and virological indicators, such as gender, age, HBV DNA level, hepatitis B e antigen (HBeAg), and ALT level. Their specificity for screening liver cancer in the high-risk groups can even reach more than 95%. In these systems, the virological indicators were highly weighted, which required extra cost and restricted the application of the model to HBV carriers. The main disadvantage of these scoring systems is that they were limited to HBV carriers, ignoring the liver cancer incidents of the HBsAg seronegative high risk population.

Increasing attention has been paid to the role of metabolic liver disruptions and the implications of these processes in liver cancer.<sup>37</sup> The unhealthy lifestyles and heavier metabolic burden increase the probability of liver cancer incidence.<sup>15, 38-43</sup> Recently, several liver cancer risk prediction models that attempted to assess risk in the general population have been developed in various settings.<sup>20, 44, 45</sup> These models predict individual risk for liver cancer within a specified period by using sociodemographic characteristics and clinical risk factors. They are limited by focusing purely on the mathematical accuracy of the model based on additional information from clinical workups, such as ALT and AFP, and do not assess predicted net benefit among different characteristics of the population. Although some prediction models of liver cancer risk include metabolic diseases such as obesity, diabetes, and nonalcoholic fatty liver disease,<sup>19, 46</sup> they were all established for specific ethnic groups and specific disease populations. These models are not universally applicable and have poor prediction efficacy in the Chinese population in which the risk factors for liver cancer may differ from those in Western populations.

Thus, we established a personalized risk prediction model for liver cancer based on data from the Chinese population. The model was separately established to identify high-risk individuals for both HBsAg seropositive and seronegative populations. We

focused on established predictor factors of liver cancer that are routinely available or can be easily ascertained. Hence, the model would be straightforward to implement on a wide scale in a primary care setting. In addition, our CKB-PLR model explicitly incorporates the predicted harms and benefits of clinical decisions based on the model predictions at a given risk threshold. A person whose risk, as predicted by the model, is above the threshold would be stratified as high-risk and selected for screening. A risk threshold of 2% means only those with estimated risk above 2% would be screened. As a result, nearly 30% HBsAg seropositive participants have estimated risk below 2% (median [interquartile range] age, 40.82 [37.42-44.84] years). Of which, 93.53% of these HBsAg carriers were men under 40 years old or women under 50 years old. This is consistent with AASLD (American Association of the Study of the Liver Diseases) screening guidelines for Asian populations with hepatitis B (i.e., males  $\geq 40$  years and females  $\geq 50$  years).<sup>47</sup> On the other hand, in view of the China's current HBsAg seropositive strategy (recommending liver cancer screening for all HBsAg seropositive participants),<sup>5</sup> and these HBsAg carriers (10-year predicted risk below 2%) only make up about 1.02% of the general adult population, screening for liver cancer does not impose a serious social or medical burden. Also, the natural history of chronic HBV infection and disease is complex and highly variable.<sup>48</sup> Taken together, we still recommend that these HBsAg seropositive participants be regularly tested for HBV virology and liver function indicators, and be screened for liver cancer. In addition, one of the strength is that our study have made up for the deficiency of liver cancer screening recommendations for HBsAg seronegative participants. As shown in **Supplementary Table 5**, HBsAg seronegative participants were divided into normal and liver cancer high-risk groups according to a probability cutoff of 10-year absolute risk of 2%, and the distribution of sociodemographic characteristics, medical history, metabolic risk factors and lifestyle factors were significantly different between the two groups. The high predictive power of factors was closely associated with metabolic diseases for liver cancer risk. Of note, as growing evidence links cholelithiasis, metabolic syndrome and liver cancer,<sup>40, 49</sup> we have incorporated a history of gallstone or gallbladder disease into our CKB-PLR model for the first time. As BMI showed an inverse association with liver cancer, our previous study performed a detailed analysis and discussion,<sup>50</sup> which appeared to be driven chiefly by reverse causation, and the discrepant causes of liver cancer may explain the different associations of BMI with liver cancer in East Asia and North America or Europe. This

did not affect BMI as a predictor to stratify participants as high-risk of liver cancer occurrence in the Chinese population. The authors specified that the calculation of the 10-year absolute risk of liver cancer should be considered only as a guide for helping to stratify the Chinese population into risk categories on a wide scale in a primary care setting.

The strengths of our study include a prospective design, a large and diverse study population, and comprehensive measurements of risk factors (e.g., HBsAg, random blood glucose, and physical activity) for liver cancer. In particular, we ascertained liver diseases through linkage to hospitalization records in addition to death and cancer registries. The model only requires information that is readily available to community health service staff and is readily amenable to further external validation. This is of crucial importance if individual risk assessment becomes an integral part of any population screening programs for liver cancer.

Some limitations of our study must be recognized. The results of the internal validation provide promising evidence that the model will perform well when applied to new populations. However, in practice, performance may vary, especially in populations with different average risks. Some important predictors, such as family history of liver cancer, detailed information on cirrhosis or fibrosis tests, treatment for HBV infection and history of nonalcoholic fatty liver disease, were not included in our baseline questionnaire.

In conclusion, we developed and internally validated a liver cancer risk prediction model using data from a large prospective cohort of the Chinese population. The model was capable of discriminating well between those at high and low risk of liver cancer, and the model-predicted probabilities were well calibrated. The clinical utility assessment of the CKB-PLR model showed potential for improved risk counseling. Our model could thus be easily implemented as an aid to physicians and individuals who may be considering undergoing screening for liver cancer.

**Acknowledgment:** The authors would like to thank all the study participants and members of the survey teams in each of the ten study regions, as well as the project development and management teams based in Beijing, Oxford, and the survey regions.

**Specific author contributions:** ZH, LL, and HS contributed to the study design and supervised the whole project. Chengxiao Yu contributed to the data interpretation, data analysis, and writing of the manuscript. CS, and JL contributed to the study design, data collection, and data interpretation of the present analysis. MZ, Canqing Yu, YG, LY, YC, ZC, TJ, HM, and GJ contributed to the study design and sample collection. All of the authors reviewed or revised the manuscript.

**Financial support:** This study was supported by National Natural Science Foundation of China (81903382, 91846303), Natural Science Foundation of Jiangsu Province (BK20190652), grants from the National Key R&D Program of China (2016YFC0900500, 2016YFC0900501, 2016YFC0900504), China Postdoctoral Science Foundation (General Program, 2019M651900). The CKB baseline survey and the first re-survey were supported by a grant from the Kadoorie Charitable Foundation in Hong Kong. The long-term follow-up is supported by grants from the UK Wellcome Trust (212946/Z/18/Z, 202922/Z/16/Z, 104085/Z/14/Z, 088158/Z/09/Z), National Natural Science Foundation of China (81390540), and Chinese Ministry of Science and Technology (2011BAI09B01).

**Competing interests:** The authors have declared that no competing interests exist.

**Ethics statement:** All participants provided informed consent. This study was approved by the ethical review committee of the Chinese Center for Disease Control and Prevention (Beijing, China) and the Oxford Tropical Research Ethics Committee, University of Oxford (UK).

**Data availability statement:** The access policy and procedures are available at [www.ckbiobank.org](http://www.ckbiobank.org). Further information is available from the corresponding author upon request.

## Reference

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* 2018;**68**: 394-424.
2. McGlynn KA, Petrick JL, London WT. Global epidemiology of hepatocellular carcinoma: an emphasis on demographic and regional variability. *Clinics in liver disease* 2015;**19**: 223-38.
3. Ji M, Liu Z, Chang ET, Yu X, Wu B, Deng L, Feng Q, Wei K, Liang X, Lian S, Quan W, Wang P, et al. Mass screening for liver cancer: results from a demonstration screening project in Zhongshan City, China. *Scientific reports* 2018;**8**: 12787.
4. Llovet JM, Bustamante J, Castells A, Vilana R, Ayuso Mdel C, Sala M, Bru C, Rodes J, Bruix J. Natural history of untreated nonsurgical hepatocellular carcinoma: rationale for the design and evaluation of therapeutic trials. *Hepatology* 1999;**29**: 62-7.
5. El-Serag HB, Davila JA. Surveillance for hepatocellular carcinoma: in whom and how? *Therapeutic advances in gastroenterology* 2011;**4**: 5-10.
6. Omata M, Cheng AL, Kokudo N, Kudo M, Lee JM, Jia J, Tateishi R, Han KH, Chawla YK, Shiina S, Jafri W, Payawal DA, et al. Asia-Pacific clinical practice guidelines on the management of hepatocellular carcinoma: a 2017 update. *Hepatology international* 2017;**11**: 317-70.
7. Arbuthnot P, Kew M. Hepatitis B virus and hepatocellular carcinoma. *International journal of experimental pathology* 2001;**82**: 77-100.
8. Liao X, Liang Z. Strategy vaccination against Hepatitis B in China. *Human vaccines & immunotherapeutics* 2015;**11**: 1534-9.
9. Liu J, Liang W, Jing W, Liu M. Countdown to 2030: eliminating hepatitis B disease, China. *Bulletin of the World Health Organization* 2019;**97**: 230-8.
10. Liu J, Zhang S, Wang Q, Shen H, Zhang M, Zhang Y, Yan D, Liu M. Seroepidemiology of hepatitis B virus infection in 2 million men aged 21-49 years in rural China: a population-based, cross-sectional study. *The Lancet Infectious diseases* 2016;**16**: 80-6.
11. Cui F, Shen L, Li L, Wang H, Wang F, Bi S, Liu J, Zhang G, Wang F, Zheng H, Sun X, Miao N, et al. Prevention of Chronic Hepatitis B after 3 Decades of Escalating Vaccination Policy, China. *Emerging infectious diseases* 2017;**23**: 765-72.
12. Liang X, Bi S, Yang W, Wang L, Cui G, Cui F, Zhang Y, Liu J, Gong X, Chen Y, Wang F, Zheng H, et al. Epidemiological serosurvey of hepatitis B in China--declining HBV prevalence due to hepatitis B vaccination. *Vaccine* 2009;**27**: 6550-7.
13. Wang FS, Fan JG, Zhang Z, Gao B, Wang HY. The global burden of liver disease: the major impact of China. *Hepatology* 2014;**60**: 2099-108.
14. Polaris Observatory C. Global prevalence, treatment, and prevention of hepatitis B virus infection in 2016: a modelling study. *The lancet Gastroenterology & hepatology* 2018;**3**: 383-403.
15. Pang Y, Kartsonaki C, Turnbull I, Guo Y, Clarke R, Chen Y, Bragg F, Yang L, Bian Z, Millwood IY, Hao J, Han X, et al. Diabetes, Plasma Glucose, and Incidence of Fatty Liver, Cirrhosis, and Liver Cancer: A Prospective Study of 0.5 Million People. *Hepatology* 2018;**68**: 1308-18.
16. Wang C, Wang X, Gong G, Ben Q, Qiu W, Chen Y, Li G, Wang L. Increased risk of hepatocellular carcinoma in patients with diabetes mellitus: a systematic review and meta-analysis of cohort studies. *International journal of cancer* 2012;**130**: 1639-48.
17. Campbell PT, Newton CC, Freedman ND, Koshiol J, Alavanja MC, Beane Freeman LE, Buring JE, Chan AT, Chong DQ, Datta M, Gaudet MM, Gaziano JM, et al. Body Mass Index, Waist Circumference, Diabetes, and Risk of Liver Cancer for U.S. Adults. *Cancer research* 2016;**76**: 6076-83.
18. Yang HI, Yuen MF, Chan HL, Han KH, Chen PJ, Kim DY, Ahn SH, Chen CJ, Wong VW, Seto WK, Group R-BW. Risk estimation for hepatocellular carcinoma in chronic hepatitis B (REACH-B): development and validation of a predictive score. *The Lancet Oncology* 2011;**12**: 568-74.
19. Rau HH, Hsu CY, Lin YA, Atique S, Fuad A, Wei LM, Hsu MH. Development of a web-based liver cancer prediction model for type II diabetes patients by using an artificial neural network. *Computer methods and programs in biomedicine* 2016;**125**: 58-65.
20. Wen CP, Lin J, Yang YC, Tsai MK, Tsao CK, Etzel C, Huang M, Hsu CY, Ye Y, Mishra L, Hawk E, Wu X. Hepatocellular carcinoma risk prediction model for the general population: the predictive power of transaminases. *Journal of the National Cancer Institute* 2012;**104**: 1599-611.
21. Chen Z, Chen J, Collins R, Guo Y, Peto R, Wu F, Li L, China Kadoorie Biobank collaborative g. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term

follow-up. *International journal of epidemiology* 2011;**40**: 1652-66.

22. Millwood IY, Li L, Smith M, Guo Y, Yang L, Bian Z, Lewington S, Whitlock G, Sherliker P, Collins R, Chen J, Peto R, et al. Alcohol consumption in 0.5 million people from 10 diverse regions of China: prevalence, patterns and socio-demographic and health-related correlates. *International journal of epidemiology* 2013;**42**: 816-27.

23. Yang G, Rao C, Ma J, Wang L, Wan X, Dubrovsky G, Lopez AD. Validation of verbal autopsy procedures for adult deaths in China. *International journal of epidemiology* 2006;**35**: 741-8.

24. Duncan MS, Freiberg MS, Greevy RA, Jr., Kundu S, Vasan RS, Tindle HA. Association of Smoking Cessation With Subsequent Risk of Cardiovascular Disease. *Jama* 2019;**322**: 642-50.

25. Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in medicine* 2002;**21**: 2175-97.

26. Durrleman S, Simon R. Flexible regression models with cubic splines. *Statistics in medicine* 1989;**8**: 551-61.

27. Muller DC, Johansson M, Brennan P. Lung Cancer Risk Prediction Model Incorporating Lung Function: Development and Validation in the UK Biobank Prospective Cohort Study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2017;**35**: 861-9.

28. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Medical decision making : an international journal of the Society for Medical Decision Making* 2006;**26**: 565-74.

29. Kerr KF, Brown MD, Zhu K, Janes H. Assessing the Clinical Impact of Risk Prediction Models With Decision Curves: Guidance for Correct Interpretation and Appropriate Use. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2016;**34**: 2534-40.

30. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC medical informatics and decision making* 2008;**8**: 53.

31. Tammemagi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, Chaturvedi AK, Silvestri GA, Riley TL, Commins J, Berg CD. Selection criteria for lung-cancer screening. *The New England journal of medicine* 2013;**368**: 728-36.

32. Pharoah PD, Antoniou AC, Easton DF, Ponder BA. Polygenes, risk prediction, and targeted prevention of breast cancer. *The New England journal of medicine* 2008;**358**: 2796-803.

33. Bruix J, Sherman M, American Association for the Study of Liver D. Management of hepatocellular carcinoma: an update. *Hepatology* 2011;**53**: 1020-2.

34. Lee HW, Ahn SH. Prediction models of hepatocellular carcinoma development in chronic hepatitis B patients. *World journal of gastroenterology* 2016;**22**: 8314-21.

35. Wong VW, Chan SL, Mo F, Chan TC, Loong HH, Wong GL, Lui YY, Chan AT, Sung JJ, Yeo W, Chan HL, Mok TS. Clinical scoring system to predict hepatocellular carcinoma in chronic hepatitis B carriers. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2010;**28**: 1660-5.

36. Yuen MF, Tanaka Y, Fong DY, Fung J, Wong DK, Yuen JC, But DY, Chan AO, Wong BC, Mizokami M, Lai CL. Independent risk factors and predictive score for the development of hepatocellular carcinoma in chronic hepatitis B. *Journal of hepatology* 2009;**50**: 80-8.

37. Satriano L, Lewinska M, Rodrigues PM, Banales JM, Andersen JB. Metabolic rearrangements in primary liver cancers: cause and consequences. *Nature reviews Gastroenterology & hepatology* 2019;**16**: 748-66.

38. Inoue M, Kurahashi N, Iwasaki M, Tanaka Y, Mizokami M, Noda M, Tsugane S, Japan Public Health Center-based Prospective Study G. Metabolic factors and subsequent risk of hepatocellular carcinoma by hepatitis virus infection status: a large-scale population-based cohort study of Japanese men and women (JPHC Study Cohort II). *Cancer causes & control : CCC* 2009;**20**: 741-50.

39. Chen CL, Yang HI, Yang WS, Liu CJ, Chen PJ, You SL, Wang LY, Sun CA, Lu SN, Chen DS, Chen CJ. Metabolic factors and risk of hepatocellular carcinoma by chronic hepatitis B/C infection: a follow-up study in Taiwan. *Gastroenterology* 2008;**135**: 111-21.

40. Vogtmann E, Shu XO, Li HL, Chow WH, Yang G, Ji BT, Cai H, Yu C, Gao YT, Zheng W, Xiang YB. Cholelithiasis and the risk of liver cancer: results from cohort studies of 134,546 Chinese men and women. *Journal of epidemiology and community health* 2014;**68**: 565-70.

41. Turati F, Galeone C, Rota M, Pelucchi C, Negri E, Bagnardi V, Corrao G, Boffetta P, La Vecchia C. Alcohol and liver cancer: a systematic review and meta-analysis of prospective studies. *Annals of*



*oncology : official journal of the European Society for Medical Oncology* 2014;**25**: 1526-35.

42. Lee YC, Cohet C, Yang YC, Stayner L, Hashibe M, Straif K. Meta-analysis of epidemiologic studies on cigarette smoking and liver cancer. *International journal of epidemiology* 2009;**38**: 1497-511.

43. Arem H, Loftfield E, Saint-Maurice PF, Freedman ND, Matthews CE. Physical activity across the lifespan and liver cancer incidence in the NIH-AARP Diet and Health Study cohort. *Cancer medicine* 2018;**7**: 1450-7.

44. Hung YC, Lin CL, Liu CJ, Hung H, Lin SM, Lee SD, Chen PJ, Chuang SC, Yu MW. Development of risk scoring system for stratifying population for hepatocellular carcinoma screening. *Hepatology* 2015;**61**: 1934-44.

45. Michikawa T, Inoue M, Sawada N, Iwasaki M, Tanaka Y, Shimazu T, Sasazuki S, Yamaji T, Mizokami M, Tsugane S, Japan Public Health Center-based Prospective Study G. Development of a prediction model for 10-year risk of hepatocellular carcinoma in middle-aged Japanese: the Japan Public Health Center-based Prospective Study Cohort II. *Preventive medicine* 2012;**55**: 137-43.

46. Ioannou GN, Green P, Kerr KF, Berry K. Models estimating risk of hepatocellular carcinoma in patients with alcohol or NAFLD-related cirrhosis for risk stratification. *Journal of hepatology* 2019;**71**: 523-33.

47. Heimbach JK, Kulik LM, Finn RS, Sirlin CB, Abecassis MM, Roberts LR, Zhu AX, Murad MH, Marrero JA. AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology* 2018;**67**: 358-80.

48. Fattovich G, Bortolotti F, Donato F. Natural history of chronic hepatitis B: special emphasis on disease progression and prognostic factors. *Journal of hepatology* 2008;**48**: 335-52.

49. Shaffer EA. Gallstone disease: Epidemiology of gallbladder stone disease. *Best practice & research Clinical gastroenterology* 2006;**20**: 981-96.

50. Pang Y, Kartsonaki C, Turnbull I, Guo Y, Chen Y, Clarke R, Bian Z, Bragg F, Millwood IY, Yang L, Huang Y, Yang Y, et al. Adiposity in relation to risks of fatty liver, cirrhosis and liver cancer: a prospective study of 0.5 million Chinese adults. *Scientific reports* 2019;**9**: 785.

## Figure Legends

**Figure 1. Projected distribution of the 10-year absolute risk for liver cancer of the Chinese population in the validation dataset.**

**Figure 2. The calibration and discrimination of the CKB-PLR (Prediction for Liver Cancer Risk based on the CKB Study) model in the validation dataset. (A)** The observed 10-year probability of liver cancer was estimated by the Kaplan-Meier method within a tenth of the model-predicted probability. **(B)** Receiver operating characteristic curve (concordance (c)-statistic: 0.80, 95% CI: 0.78-0.82). **(C)** Sensitivity and specificity on the basis of the predicted probability cutoff.

**Figure 3. Decision curves of the CKB-PLR (Prediction for Liver Cancer Risk based on the CKB Study) model compared with HBsAg seropositive criteria\* for liver cancer screening decisions. \*HBsAg Seropositive criteria were defined as people with hepatitis B virus infection.**

**Table 1. Distribution of demographic characteristics and covariates.**

Characteristics	No Liver Cancer (N=483,579)		Liver Cancer (N=2,706)		Incidence Rate Per 100,000 person years (95% CI)*
	No.	%	No.	%	
<b>Gender</b>					
Female	286,981	59.35	967	35.74	29.56 (27.67-31.57)
Male	196,598	40.65	1,739	64.26	73.66 (69.85-77.67)
<b>Age at baseline (Years)</b>					
Mean age (years, SD)	51.80 ± 10.61		58.12 ± 9.99		
<50	221,349	45.77	578	21.36	23.49 (21.62-25.53)
50-59	148,853	30.78	931	34.41	56.57 (52.93-60.46)
60-69	83,687	17.31	841	31.08	92.21 (85.93-98.94)
≥70	29,690	6.14	356	13.16	117.46 (105.58-130.67)
<b>Residential area</b>					
Rural	267,309	55.28	1,523	56.28	45.11 (42.62-47.76)
Urban	216,270	44.72	1,183	43.72	40.30 (37.78-42.99)
<b>Highest education</b>					
Primary School and lower	241,733	49.99	1,629	60.20	46.83 (44.00-49.84)
Middle school and higher	241,846	50.01	1,077	39.80	38.95 (36.41-41.66)
<b>Family history of cancer</b>					
No	399,675	82.65	2,164	79.97	41.61 (39.54-43.78)
Yes	83,904	17.35	542	20.03	48.79 (44.65-53.31)
<b>Previous cancer diagnosis</b>					
No	481,207	99.51	2,665	98.48	42.56 (40.60-44.62)
Yes	2,372	0.49	41	1.52	118.46 (86.98-161.34)
<b>History of diabetes</b>					
No	455,060	94.10	2,427	89.69	41.56 (39.59-43.62)
Yes	28,519	5.90	279	10.31	63.97 (56.51-72.42)
<b>History of cirrhosis or chronic hepatitis</b>					
No	478,148	98.88	2,455	90.72	39.71 (37.83-41.68)
Yes	5,431	1.12	251	9.28	315.88 (277.85-359.10)
<b>History of gallstone or gallbladder disease</b>					
No	454,700	94.03	2,402	88.77	40.23 (38.29-42.26)
Yes	28,879	5.97	304	11.23	83.20 (74.06-93.45)
<b>Body-mass index (kg/m<sup>2</sup>)</b>					
<18.5	20,540	4.25	159	5.88	50.06 (42.59-58.83)
18.5-23.9	249,884	51.67	1,465	54.14	44.57 (42.00-47.31)
24-27.9	161,454	33.39	821	30.34	39.89 (37.06-42.93)
≥28	51,701	10.69	261	9.65	42.07 (37.17-47.62)
<b>Hepatitis B test</b>					

---

Seronegative	469,098	97.01	2,019	74.61	31.59 (29.95-33.32)
Seropositive	14,481	2.99	687	25.39	401.96 (371.64-434.76)
<b>Random plasma glucose (mmol/L)</b>					
<5.6	235,554	48.71	1,133	41.87	38.72 (36.32-41.28)
5.6-6.7	154,402	31.93	838	30.97	42.47 (39.46-45.70)
6.8-7.7	50,903	10.53	351	12.97	49.42 (44.29-55.15)
7.8-11.0	28,177	5.83	231	8.54	54.47 (47.63-62.30)
≥11.1	14,543	3.01	153	5.65	70.74 (60.09-83.27)
<b>Physical activity (MET-hours/day)</b>					
<10.32	120,019	24.82	985	36.40	47.51 (43.88-51.43)
10.32-17.46	121,332	25.09	647	23.91	43.34 (39.94-47.02)
17.47-30.00	121,062	25.03	586	21.66	43.61 (40.10-47.42)
≥30.01	121,166	25.06	488	18.03	37.62 (34.29-41.27)
<b>Tobacco smoking</b>					
Never	328,797	67.99	1,268	46.86	38.39 (36.13-40.79)
Smoker	154,782	32.01	1,438	53.14	53.01 (48.80-57.58)
<b>Alcohol drinking</b>					
Never	393,232	81.32	1,850	68.37	40.83 (38.79-42.97)
Drinker	90,347	18.68	856	31.63	51.81 (47.61-56.39)

---

\* Incidence rate for liver cancer are adjusted for age at baseline, gender, and residential area.

**Table 2. Detailed descriptions of the model parameters in HBsAg seronegative and seropositive participants in the development dataset.**

Predictor	HBsAg Seronegative		HBsAg Seropositive	
	HR (95% CI)*	Beta	HR (95% CI)*	Beta
First spline basis of age (knots:30, 50, 60, 70, 80)	5.72 (5.08-6.43)	1.744	10.91 (8.98-13.16)	2.390
Second spline basis of age (knots:30, 50, 60, 70, 80)	8.81 (7.61-10.16)	2.176	2.99 (2.09-4.19)	1.094
Third spline basis of age (knots:30, 50, 60, 70, 80)	26.92 (22.71-31.80)	3.293	62.93 (48.62-80.72)	4.142
Forth spline basis of age (knots:30, 50, 60, 70, 80)	4.65 (3.71-5.84)	1.537	14.89 (10.33-21.67)	2.700
Male (vs.female)	1.78 (1.68-1.90)	0.579	2.73 (2.47-3.00)	1.003
Residential area (urban vs. rural)	0.88 (0.82-0.95)	-0.125	0.83 (0.74-0.94)	-0.183
Highest education (middle school and higher vs. others)	0.79 (0.72-0.85)	-0.241	0.99 (0.88-1.11)	-0.009
Family history of cancer (yes vs. no)	1.02 (0.91-1.15)	0.024	1.26 (1.06-1.47)	0.228
Previous cancer diagnosis (yes vs. no)	2.55 (1.69-3.66)	0.936	1.60 (0.64-3.24)	0.470
History of diabetes (yes vs. no)	1.20 (1.02-1.38)	0.179	1.18 (0.90-1.51)	0.167
History of cirrhosis or chronic hepatitis (yes vs. no)	5.41 (4.37-6.59)	1.687	2.00 (1.65-2.39)	0.692
History of gallstone or gallbladder disease (yes vs. no)	1.99 (1.72-2.28)	0.687	1.34 (1.00-1.77)	0.296
Body-mass index, per 1-unit increase	0.97 (0.97-0.98)	-0.025	0.98 (0.98-0.99)	-0.017
Physical activity, per 1-unit increase	0.99 (0.99-0.99)	-0.008	1.00 (0.99-1.00)	-0.005
First spline basis of random glucose (knots: 2.8, 5.6, 6.8, 11.1)	1.28 (1.19-1.38)	0.246	1.81 (1.63-2.06)	0.594
Second spline basis of random glucose (knots: 2.8, 5.6, 6.8, 11.1)	1.32 (1.20-1.45)	0.276	2.88 (2.44-3.37)	1.056
Third spline basis of random glucose (knots: 2.8, 5.6, 6.8, 11.1)	1.34 (1.25-1.43)	0.296	1.82 (1.62-2.02)	0.600
Tobacco regular consumption (smokers vs. never)	1.21 (1.13-1.30)	0.192	1.33 (1.19-1.48)	0.284
Alcohol regular consumption (drinkers vs. never)	1.40 (1.29-1.53)	0.339	1.03 (0.89-1.19)	0.033

\* HR, hazard ratio. CI, confidence interval.

**Table 3. Sensitivity, Specificity, and Net Benefit at risk threshold of 2% of the 10-year CKB-PLR (Prediction for Liver cancer Risk based on the CKB study) model.**

Criteria	Sensitivity, %	Specificity, %	Net benefit (95% CI)*
<b>HBsAg Seropositive<sup>†</sup></b>	25.96%	96.96%	0.000734 (0.000623-0.000826)
<b>CKB-PLR<sup>‡</sup></b>	28.17%	97.12%	0.000912 (0.000805-0.001022)

<sup>†</sup>HBsAg Seropositive criteria were defined as people with hepatitis B virus infection;

<sup>‡</sup>According to the CKB-PLR (Prediction for Liver cancer Risk based on the CKB study) model criteria, positivity (eligible for screening) was defined as a probability of liver cancer greater than 2% over a period of 10 years;

\* Net benefit at threshold of 2% predicted absolute risk.